

Internal interannual variability of the winter polar vortex in a simple model of the seasonally evolving stratosphere

L. A. Hatfield and R. K. Scott*

School of Mathematics and Statistics, University of St Andrews, UK

*Correspondence to: School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews, KY16 9SS, UK. E-mail: rks4@st-andrews.ac.uk

We investigate persistent low-frequency variability of the stratospheric winter polar vortex in a rotating spherical shallow-water model under the action of topographic wave-forcing and radiative cooling to a simple time-varying equilibrium state representative of the seasonal cycle in solar heating. A range of modes of variability is obtained, dependent on wave forcing amplitude and characterized by the distribution of quiescent and disturbed winters, defined as winters in which the vortex is either close to radiative equilibrium, with low planetary wave amplitude, or else strongly disturbed from equilibrium by the wave forcing. At low forcing amplitude the vortex is typically quiescent every year, while at higher amplitude it is typically disturbed; in both cases there is little year-to-year variation of the vortex state. For a range of intermediate forcing amplitudes, however, the vortex transitions between quiescent and disturbed states from one winter to the next with a persistent and well-defined pattern of variability. To investigate the extent to which the low-frequency variability found here may be explained in terms of a low-latitude flywheel mechanism, we conduct additional experiments varying a linear drag on the zonal mean flow in the tropics and find that sufficiently strong drag can completely suppress the variability. The robustness of the variability is demonstrated by further experiments using a modified radiative equilibrium profile, associated with a tropical westerly flow: similar variability is obtained but the modified profile is less effective at constraining the tropical flow from a persistent easterly acceleration. Copyright © 2013 Royal Meteorological Society

Key Words: polar vortex; low-frequency variability; shallow-water model; seasonal cycle

Received ...

Citation: ...

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.3604

1. Introduction

Owing to its strong density stratification and low moisture content, the Earth's stratosphere is one of the most dynamically stable parts of the atmosphere, yet possesses distinct variability forced by upward propagating planetary waves. Its dynamics is closely coupled to that of the underlying troposphere through long range wave propagation and eddy transports of heat and momentum, and consequently has an impact on surface weather patterns and climate (e.g. Haynes 2005; Gerber and Polvani 2009). Understanding its variability, on both short and long timescales, is therefore highly relevant for both climate prediction and seasonal forecasting.

One of the most significant modes of stratospheric variability is the year-to-year variation of the winter polar vortex. This manifests itself most obviously in the Northern hemisphere, where the strength of wave forcing from topography and land-sea temperature contrasts is such that the frequency of stratospheric sudden warming events, where the vortex is strongly disturbed from radiative equilibrium, is approximately one every other year. Such events can have a large impact on the troposphere below (Thompson and Wallace (1998), Baldwin and Dunkerton (1999), Baldwin and Dunkerton (2001), Thompson *et al.* (2002)), the intense cold period in Europe in early 2018 being just one recent example of their effects on surface weather.

The conditions that determine whether a sudden warming will occur in a given year are complex and not fully understood, but we may loosely regard the sources of variability of the polar vortex as being either *external* or *internal*. External sources may include phenomena such as the El Niño-Southern Oscillation (ENSO) or the solar cycle, which can be considered as variability in the dynamical or radiative forcings of the stratosphere. The Quasi-Biennial Oscillation (QBO), although existing in the stratosphere, could also be regarded as a source of variability external to the midlatitude dynamics and the polar vortex.

There is also strong evidence for the existence of internal dynamical mechanisms in the winter stratosphere that may give rise to variability in the polar vortex even in the absence of any variability in the external forcings or boundary conditions. Understanding such internal modes is highly relevant for studies into the variability of the rest of the atmosphere, as they imply that the stratosphere itself can have a direct influence on the complete atmospheric circulation, rather than simply acting as a mediator to it.

Determining the nature of and mechanisms behind such internal variability is difficult due to the complexity of the dynamics. Modelling the stratosphere more accurately and in more detail can improve forecasts but is computationally expensive and may be unfeasible when investigating long-term trends. Causal mechanisms can also be more difficult to identify in a system with a larger number of interacting variables. An alternative approach with a long history of success is the use of simpler, dynamically reduced models. When investigating internal variability, a simpler model has the obvious advantage in that all external sources of variability can be precisely controlled or removed altogether, enabling internal processes to be examined in isolation. Further, when investigating low-frequency variability, the simple model has the practical advantage of being computationally efficient, allowing runs of sufficient duration needed to establish long-term patterns.

The pioneering study of Holton and Mass (1976) was the first to suggest that the winter stratosphere exhibited internal modes of variability. Using a perpetual January quasi-geostrophic beta-plane model, they found that with sufficiently large topographic forcing amplitude the zonal winds at 60° N oscillated quasi-regularly between easterlies and westerlies in a manner characteristic of stratospheric sudden warming events. The particular significance of this study and subsequent studies using similar reduced models (Holton and Dunkerton (1978), Chao (1985), Schoeberl and Strobel (1980), Yoden (1987)) is that the topographic forcing is time independent, and hence any periodicity in the flow is the result of an internal mode of variability.

Accepted Article

These studies were extended by Christiansen (1999), Scott and Haynes (2000), Scott and Polvani (2006) using more realistic primitive equation models and less severe spectral truncation, the latter providing a comprehensive study into internal polar vortex vacillations at a reasonably high, horizontally isotropic resolution. They found consistent variability, again in the form of regular vacillations of the polar vortex on similar timescales of approximately 50–100 days across a wide range of topographic forcing amplitudes and radiative forcing strengths. Patterns of stratospheric variability have also been identified in more comprehensive models with an active troposphere (e.g. Taguchi *et al.* 2001), although their interpretation is complicated by the sensitivity of tropospheric dynamical time scales to model resolution (Gerber *et al.* 2008).

In stratospheric primitive equation models, the vertical structure of the polar vortex has appeared an important ingredient of the vacillation cycles. The destruction of the vortex generally begins at the upper levels and propagates downwards, generating a large region of weak easterlies which subsequently act as a barrier to further upward wave propagation. This allows the vortex to reform first at the upper levels under radiative cooling and initiate another cycle.

More recently, it has become recognized that much vortex variability, at least that involving planetary wavenumber two, is associated with the excitation of a barotropic mode, with relatively constant-in-height vertical structure (Esler and Scott 2005; Matthewman *et al.* 2009). The extent to which persistent internal variability may arise in the absence of vertical structure has also recently been addressed. Rong and Waugh (2004) conducted an early study on internal variability of the polar vortex using a shallow-water model under perpetual January conditions. The absence of vertical structure in the shallow-water system means that any variability obtained can be attributed to purely barotropic dynamics and cannot be due to details of the vertical structure of the vortex.

Instead, Rong and Waugh (2004) proposed that the vacillation cycles in their model were due to a self-perpetuating cycle involving the formation of strong potential vorticity gradients on the vortex edge. These act as a wave guide for Rossby waves, which eventually break, mixing potential vorticity to small scales where it is dissipated, weakening potential vorticity gradients and Rossby wave activity, and thus allowing the vortex to reform under radiative effects. The vacillations obtained by Rong and Waugh (2004), however, were not persistent beyond a few cycles and, moreover, were sensitive to increasing horizontal resolution. MirRokni *et al.* (2011) performed a more detailed analysis of the shallow-water system using two alternative numerical techniques and found dynamically similar vacillations that again diminished rapidly and vanished altogether after only a few cycles.

A feature of the model simulations of both Rong and Waugh (2004) and MirRokni *et al.* (2011) was a persistent growth of easterlies in the tropics, which arises spontaneously in the spherical shallow-water system by Rossby wave breaking in the lower latitudes transferring negative angular pseudomomentum into the tropics. The persistent transfer of momentum is constrained only weakly at low latitudes by thermal relaxation and in time easterlies increase in the tropics until they eventually infringe on the dynamics of the winter polar vortex, preventing the variability from continuing indefinitely. As a solution, Scott (2016) introduced a weak drag on the zonal flow in the tropics to explicitly control the unwanted build up of easterly winds. The drag was restricted to the tropics to avoid artificially damping the midlatitude dynamics and the polar vortex. With this modification, the polar vortex exhibited vacillations over extended time intervals, again with periods of the order of 100 days. The conclusion was that the shallow water system does indeed possess persistent internal variability, provided the tropical easterlies are controlled, suggesting in turn that there is an internal mode of stratospheric variability involving only the barotropic mode and not reliant on details of the vertical structure.

Accepted Article

In light of the above, a natural question that now arises is how the variability of the polar vortex in the shallow water system survives, or is modified by, the inclusion of a seasonal cycle in the radiative forcing. Instead of considering persistent, short-timescale vacillations of the polar vortex in perpetual January, are there clear modes of internal, interannual variability in the presence of a seasonal cycle? The purpose of the current paper is to address this question. Specifically, we extend the above studies with the introduction of a seasonal cycle in the radiative forcing, while keeping the wave forcing constant in time. Such a radiative forcing, through thermal wind balance, produces a seasonally strengthening and diminishing polar vortex as the polar regions are alternately cooled and heated, and we therefore consider internal interannual variability as that which results in a clear pattern of year-to-year variations in the vortex strength.

A mechanism for internal interannual variability in the atmosphere, involving the idea that low-latitudes can act as a “sub-tropical flywheel” for angular momentum anomalies, was proposed by Scott and Haynes (1998). Radiative relaxation has a relatively strong effect on the zonal flow in midlatitudes: dynamical anomalies generated by wintertime wave breaking are damped on time scales much shorter than the seasonal one, and so flow anomalies do not easily persist from one winter to the next, and cannot influence a subsequent winter’s evolution. The summer stratosphere is a dynamically inactive period and can be considered to act as an eraser of winter flow anomalies. At lower latitudes, however, thermal wind balance is a much looser constraint between the dynamical and thermal fields: radiative relaxation consequently has a weaker effect on dynamical anomalies, which may persist for longer. At low enough latitudes the persistence times may be such that dynamical anomalies from one winter retain sufficient magnitude by the onset of the following winter to have an influence on the subsequent evolution. The effect may also be considered in terms of a dispersion relation for zonal mean flow anomalies (see Appendix to Scott and

Haynes 1998, the shallow water analogue being obtained by replacing N^2/m^2 with gH in their equation A.6).

As we shall see, the dynamical anomalies preserved at lower latitudes are also important for the interannual variability of the winter polar vortex in the shallow water model. As such, care is needed when adding damping or other mechanisms to control the build-up of low-latitude easterlies just discussed. If the damping extends over too broad a region, or the timescale of the damping is too short, easterlies can be controlled but the interannual variability of the system may be lost due to the suppression of the flywheel effect at low latitudes. Conversely, if the damping is not extensive or strong enough, tropical easterlies will not be sufficiently controlled and may encroach on the flow at higher latitudes such that the interannual variability of the system is again lost (see sections 2.5 and 3.3 below for more details).

The remainder of this paper is organised as follows. In section 2 we describe the setup and parameters of the model, the seasonally varying radiative forcing, the constant topographic wave forcing, and modifications to control tropical easterlies. In section 3 we present the results of our main investigations using a linear drag on the zonal flow to control the tropical easterlies and examine the nature and mechanism of the internal interannual variability obtained. In section 4 we briefly consider the variability obtained using an alternative method for controlling the tropical easterlies, involving a modification to the radiative equilibrium height profile rather than a damping on the zonal flow. In section 5 we present our summary and conclusions.

2. The Model

2.1. Physical model

We solve the equations for the rotating shallow-water system in a spherical domain,

$$\frac{\partial \zeta}{\partial t} = -\nabla \cdot (\mathbf{u} \zeta_a), \quad (1)$$

$$\frac{\partial \delta}{\partial t} = -\frac{1}{2} \nabla^2 |\mathbf{u}|^2 + \mathbf{k} \cdot \nabla \times (\mathbf{u} \zeta_a) - g \nabla^2 (h + h_{\text{tp}}), \quad (2)$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (h \mathbf{u}) = -\alpha (h - h_e), \quad (3)$$

where ζ , δ and h are the relative vorticity, divergence, and layer thickness respectively. $\zeta_a = 2\Omega \sin \phi + \zeta$ is the absolute vorticity, where Ω is the planetary rotation rate, ϕ is the latitude, g is the gravitational acceleration and $\mathbf{u} = (u, v)$ is the horizontal velocity. The system is forced by specified wavenumber-two topography h_{tp} and a relaxation to an equilibrium height profile $h_e(\phi, t)$ at a rate α .

Following Liu and Scott (2015), we use a mean layer depth of $H = 10\text{km}$, giving a shallow-water polar Rossby deformation length $L_D = \sqrt{gH}/2\Omega \approx 2100\text{km}$, approximately equal to the Rossby deformation length of the Lamb mode in the winter stratosphere, $ND/f\sqrt{\kappa(1-\kappa)}$, where D is a density scale height and $\kappa = 2/7$.

The numerical implementation is described in Rivier *et al.* (2002) and Scott *et al.* (2004). We use a standard pseudo-spectral model with maximum total wavenumber 85 (T85) for the main experiments, with some individual cases run at T170 to verify robustness. A weak hyperdiffusion is used to dissipate enstrophy at the smallest scales; this and other numerical details are as described in Scott (2016).

2.2. Radiative forcing

The seasonal cycle in solar insolation is implemented in the model via the height profile, which can be considered as representing the thickness of an atmospheric layer between two potential temperature surfaces. The height profile is relaxed to an equilibrium state $h_e(\phi, t)$ through

the right-hand-side term of (3), anomalously high or low h_e corresponding to a heating or cooling, respectively. To construct h_e , we define a perturbation height profile for the Northern hemisphere winter solstice, $h_w(\phi)$, and a profile for the Northern hemisphere summer solstice, $h_s(\phi)$ related by $h_s(\phi) = h_w(-\phi)$, i.e., h_s is a reflection in latitude of h_w . These profiles are then multiplied by a sinusoidal time variation and superposed, to give a seasonally varying perturbation height profile $h'_e(\phi, t)$ of the form

$$h'_e(\phi, t) = s(t)h_w(\phi) + (1 - s(t))h_s(\phi), \quad (4)$$

where $s(t) = \frac{1}{2} (1 + \cos(\frac{2\pi t}{365}))$ varies between 0 and 1 with a period of 1 year. The perturbation height profile has a mean of zero, and is added to the mean layer depth H to give the final height profile $h_e(\phi, t) = H + h'_e(\phi, t)$ used in (3).

In the experiments described below we define h_w to be a simple anti-symmetric cos-squared perturbation height profile, cooling over the Northern hemisphere winter pole, with amplitude h_0 and low-latitude extent ϕ_h , by

$$h_w(\phi) = \begin{cases} h_0 \cos^2\left(\frac{\pi}{2} \left(\frac{\phi + \frac{\pi}{2}}{-\phi_h + \frac{\pi}{2}}\right)\right) & -\frac{\pi}{2} \leq \phi \leq -\phi_h \\ 0 & -\phi_h \leq \phi \leq \phi_h \\ -h_0 \cos^2\left(\frac{\pi}{2} \left(\frac{\phi - \frac{\pi}{2}}{\phi_h - \frac{\pi}{2}}\right)\right) & \phi_h \leq \phi \leq \frac{\pi}{2}. \end{cases}$$

The low-latitude extent of the heating anomaly is set to be $\phi_h = \pi/8 = 22.5^\circ$, and the amplitude h_0 is taken to be $0.6H$. These parameters were chosen after some experimentation to give a broadly representative winter polar vortex, with a maximum jet speed of approximately 100ms^{-1} located in the vicinity of 60°N .

2.3. Topographic forcing

In the absence of forcing other than radiative relaxation to the seasonally varying height profile, the system settles rapidly into a regular, annually periodic flow. To represent the effects of planetary wave forcing from topography and land-sea temperature contrasts, we prescribe a

lower boundary height perturbation h_{tp} in the Northern Hemisphere only, with a zonal wavenumber two structure, of the form

$$h_{\text{tp}}(\phi, \lambda) = \begin{cases} h_b H \cos^2 \phi \cos 2\lambda & \phi > 0 \\ 0 & \phi < 0, \end{cases} \quad (5)$$

where h_b is a dimensionless parameter that specifies the height of the topography relative to the mean layer depth. The restriction of the topographic forcing to the Northern Hemisphere is analogous to the actual asymmetry of surface topography between the hemispheres, although our primary motivation for the asymmetry is to restrict variability to one hemisphere, simplifying the investigations. Our focus on wavenumber two is motivated by the predominantly barotropic structure of wavenumber-two “vortex splitting” stratospheric sudden warming events, and its natural representation in the shallow water model.

We perform simulations with topographic forcing amplitudes in the range $0.04 \leq h_b \leq 0.18$. Values of h_b lower than this result in minimal to no impact on the polar vortex, whilst higher values result in unphysical flow, with the polar vortex failing to form at all in most winters.

2.4. Low-latitude momentum damping

As described in the introduction, under the action of sustained wave forcing, the shallow water model tends to develop unphysically strong winds in the tropics. Because these are driven by the transfer of negative angular pseudomomentum from high to low latitudes by wave breaking, the effect is more pronounced with stronger topographic wave forcing. If unconstrained, these tropical easterlies can become extremely strong: as an example, using the height profile as defined in section 2.2 and moderate topographic forcing amplitude of $h_b = 0.1$, the equatorial winds reach speeds of up to 100 ms^{-1} after as little as 5 years. Under these conditions, the easterly zone also encroaches far enough into midlatitudes that the winter evolution is no longer characteristic of the stratosphere.

Tropical easterlies in the model can be controlled in different ways and in this investigation we tried two alternative approaches. The first and primary approach was to introduce a linear drag, or friction, on the zonal mean zonal flow across the tropics and into the Southern hemisphere.

The principal advantage of using a drag directly on the zonal flow is that the easterlies are controlled with minimal further complication of the model. The flow more closely resembles realistic stratospheric flow, and diagnosis of variability is more straightforward. We implement this drag by adding the following term to the right hand side of (1):

$$-\mathbf{k} \cdot \nabla \times (r(\phi) \bar{\mathbf{u}}, 0), \quad (6)$$

where u is the zonal component of \mathbf{u} , the bar denotes the zonal average, and where $r(\phi)$ is a latitudinally dependent drag coefficient. The latitudinal dependence of $r(\phi)$ is a simple step-like profile with a plateau over the Southern hemisphere and extending into the Northern hemisphere to a latitude ϕ_{fr} . The edge of the step is smoothed by a piecewise cos-squared function with a total width of 5 degrees, the smoothness of which ensures dynamical continuity. Explicitly, the profile is given by

$$r(\phi) = \begin{cases} \tau_{\text{fr}}^{-1} & -\frac{\pi}{2} \leq \phi \leq \phi_{\text{fr}} \\ \tau_{\text{fr}}^{-1} \cos^2(18(\phi - \phi_{\text{fr}})) & \phi_{\text{fr}} \leq \phi \leq \phi_{\text{fr}} + \frac{\pi}{36} \\ 0 & \phi_{\text{fr}} + \frac{\pi}{36} \leq \phi \leq \frac{\pi}{2}. \end{cases}$$

where τ_{fr} is the frictional timescale on the step.

We seek a magnitude τ_{fr} and latitudinal extent ϕ_{fr} of the relaxation sufficient to control the build-up of easterlies without eliminating possible dynamical signals in low latitudes that might be involved in interannual variability, the subtropical flywheel discussed in the introduction. It turns out, however, that only a relatively weak and tropically confined damping is needed to sufficiently control the easterlies. After some experimentation, a suitable latitudinal extent of the relaxation was found to be approximately $\phi_{\text{fr}} =$

15 degrees, while a long frictional timescale of one year $\tau_{fr} = 365$ days could be taken. With these values, tropical easterlies are sufficiently controlled, while dynamical anomalies are preserved in the tropics for sufficient duration to permit interannual variability.

2.5. Modified radiative forcing

A possible alternative method for controlling the build-up of tropical easterlies is through a modification of the radiative equilibrium height profile. The aim is to prevent the build-up of easterly momentum by thermal forcing to an equilibrium state whose associated velocity profile has westerlies in the tropics. This method of controlling the tropical easterlies is, from one perspective, more appealing than the frictional method, as no new forcings need to be introduced into the system. On the other hand, we know that radiative relaxation is less effective at controlling low latitude angular momentum anomalies.

In the definition of the Northern Hemisphere winter solstice perturbation height profile, $h_w(\phi)$, the flat zone across the tropics, between $\phi_h = \pm \frac{\pi}{8}$, is replaced with a local, symmetric cos-squared perturbation of amplitude $h^* = 0.06H$:

$$h_w(\phi) \longrightarrow h_w(\phi) + h^* \cos^2\left(\frac{\pi\phi}{2\phi_h}\right), \quad (7)$$

while outside the tropics, for $|\phi| > \phi_h$, h_w is unchanged. Note that because the perturbation is symmetric in ϕ it contributes only a steady part to the full $h_{eq}(\phi, t)$. The global mean perturbation height is removed to ensure the global mean of h_{eq} is unaltered by the modification.

This additional local height perturbation over the equator, is in thermal wind balance with a corresponding equilibrium zonal velocity profile comprising westerly winds in the tropics. It might therefore be expected to help maintain the tropics in a more westerly state and, with appropriate choice of strength and latitudinal extent, counteract the build-up of easterly winds there. In practice the situation is more complex and the heating anomaly is only partially able to

control the easterly flow. This will be discussed in more detail in section 4 below, where experiments using this thermal forcing are presented.

3. Results: tropical drag

In this section we present results from model integrations in which the tropical easterlies are controlled by a linear drag on the zonal velocity (section 2.5). In all of the following experiments we investigate only variability of the winter polar vortex. Our discussion is based on a categorisation of each winter as either “quiescent” or “disturbed”, and the distribution and sequence of the winter states using this categorisation. Model integrations are all of a duration of one hundred years, starting from a state of rest.

3.1. Overview of the variability

The winter of each year will be defined as the first 90 days of the year (analogous to 21st December to 21st March), since this is roughly the 3-month period during which the zonal velocity and potential vorticity are highest in the Northern hemisphere, and encompasses the bulk of the development and dissipation of the Northern winter polar vortex.

Broadly speaking, the polar vortex in a quiescent winter is characterised by a steady increase and decrease in zonal velocity and potential vorticity at high latitudes over the winter. The zonal flow is held relatively close to the radiative equilibrium and the vortex retains a regular, slightly elliptical shape. As an example, Figure 1(a) shows the last ten years of the Northern hemisphere zonal mean zonal wind as a function of latitude for the lowest forcing amplitude considered, $h_b = 0.04$. The ten year period shown is entirely representative of the complete one hundred year integration. Every winter is quiescent without exception. The time evolution of the zonal velocity field is very smooth, and closely resembles the unforced Southern hemisphere (not shown). The zonal velocity steadily increases to a maximum of approximately 90 ms^{-1} at midwinter, centered on about 60° N , then gradually decreases in magnitude over the remainder of the winter.

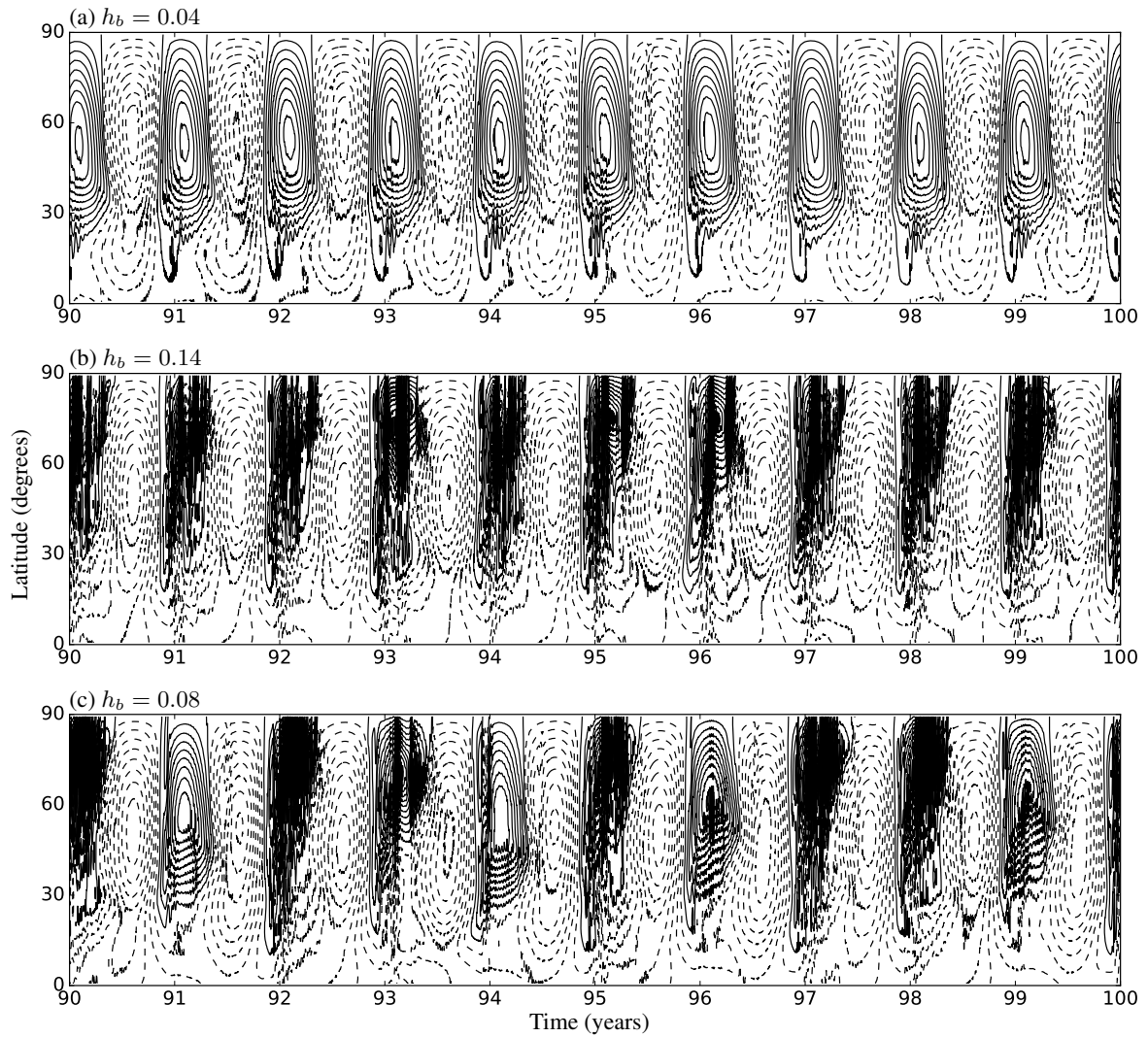


Figure 1. Zonal mean zonal velocity, \bar{u} , for the final 10 years of the 100 year time integrations for various topographic forcing amplitudes (top to bottom: $h_b = 0.04$, $h_b = 0.14$ and $h_b = 0.08$). Contour spacing = 10ms^{-1} , zero and negative contours are dashed.

In contrast to this evolution, the polar vortex in a disturbed winter is characterised by a highly irregular evolution of zonal velocity and potential vorticity at high latitudes. The zonal flow is held far from radiative equilibrium by persistent wave-breaking, weakening the jet and pushing the vortex edge farther north. The vortex itself distorts, filaments, and often splits into two, usually multiple times over the winter. As an example, Figure 1(b) shows the evolution over the same ten year period but for much higher topographic forcing amplitude, $h_b = 0.14$. In this case almost every winter is disturbed. Details of individual winters differ but each one is essentially disturbed and characterized by erratic and highly transient evolution of the zonal mean zonal velocity. On the other hand, the summer evolution remains relatively undisturbed:

as in the stratosphere, the summer easterlies prevent the topographic forcing from exciting significant wave motions and most eddy activity is damped out from one year to the next by the radiative relaxation. Note that in both examples just considered, there is very little systematic year-to-year variation of the vortex state.

The question now arises as to what state the vortex takes for topographic forcing values in between these two extremes. One could imagine that as forcing is increased gradually from $h_b = 0.04$, the evolution of each winter might become gradually more disturbed, yet with year-to-year variations remaining small. In fact, as shown in Figure 1(c), for the case $h_b = 0.08$, we see a series of winters, each of which resembles one or other of the two cases just described, i.e. there is a mixture of quiescent and

disturbed winters that are obtained for this single forcing amplitude. For example, year 94 closely resembles a typical quiescent winter from the low forcing amplitude case $h_b = 0.04$, while year 95 is very similar to a typical disturbed winter from the very high forcing amplitude case $h_b = 0.14$.

As will be seen in the following sections, the type of behaviour exhibited in Figure 1(c) is quite characteristic of a wide range of intermediate forcing amplitudes, between about $0.06 \leq h_b \leq 0.14$, that are of the primary interest in this paper. We emphasize at this point that the great majority of all winters obtained at each of these intermediate forcing amplitudes are essentially either of quiescent or disturbed type, and that there are relatively few winters of an inbetween, mildly disturbed, nature. The two years mentioned above, years 94 and 95 of the run $h_b = 0.08$ are thus typical of quiescent and disturbed winters for almost all intermediate forcing amplitudes. In a very few years the polar vortex does seem to exhibit a mixture of quiescent and disturbed behaviour within a single winter. However, a closer analysis shows that such hybrid winters are essentially disturbed winters in which the onset of the disturbance occurs unusually late in the winter season.

The number of disturbed winters in each integration increases almost monotonically with the forcing amplitude. For low forcing amplitudes below $h_b = 0.055$, every winter is quiescent, and for higher forcing amplitudes above about $h_b = 0.14$, almost every winter is disturbed. For intermediate forcing amplitudes in the range $0.07 \leq h_b \leq 0.1$, quiescent and disturbed winters occur in roughly equal frequency. The dependence of the distribution of quiescent and disturbed winters will be discussed in the next section. First we give a more complete description of the two types of evolution.

Figure 2 shows the evolution of the zonal velocity over the winter (the 90 days period beginning on the solstice) for the two consecutive years of the model integration with forcing amplitude $h_b = 0.08$. The variable plotted is the

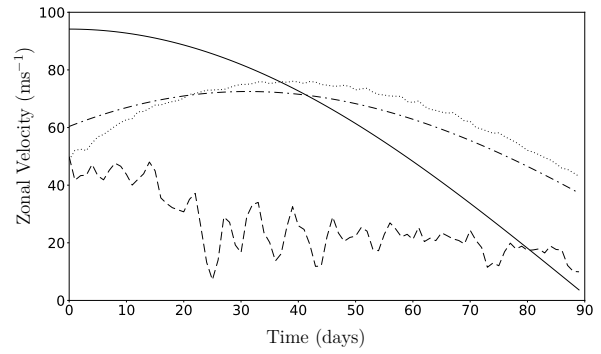


Figure 2. Latitude-weighted mean of the zonal velocity vs. time between 50° and 70° North for two consecutive winters (dotted line year 94; dashed line year 95) of the run for $h_b = 0.08$. The unforced run ($h_b = 0$, dash-dotted line) and the radiative equilibrium (solid line) are also displayed.

latitudinally-weighted mean of the zonal velocity, given by

$$\langle \bar{u} \rangle_\phi(t) = \frac{\int_{\phi_1}^{\phi_2} \bar{u} \cos \phi d\phi}{\int_{\phi_1}^{\phi_2} \cos \phi d\phi}, \quad (8)$$

with lower and upper bounds for latitude $\phi_1 = 50^\circ$ and $\phi_2 = 70^\circ$, chosen to capture the distinction between the two winter states. This is essentially a measure of the angular momentum of the strongest part of the vortex, and a natural way of quantifying the “strength” of the vortex at a time t . The difference between the two years is clear: year 94, the quiescent year (dotted line), exhibits a smooth, almost perfectly monotonic increase and subsequent decrease of the vortex strength; it follows closely the evolution of an integration in which the wave forcing is absent (dash-dotted line), the small differences being due to a slight contraction of the vortex in latitude and shift in the location of the jet maximum. It also follows the time varying radiative equilibrium state (solid line) with a lag due to the finite radiative relaxation timescale. In contrast, in year 95, the disturbed year, the evolution of the vortex is much more erratic and non-monotonic; it is held much farther from radiative equilibrium, and remains significantly weaker than the vortex of the quiescent year for the duration of the winter.

The evolution of the zonal mean zonal velocity gives a broad indication of the strength of the vortex but does not provide information about the vortex structure and other details of its evolution, in particular the nature of the

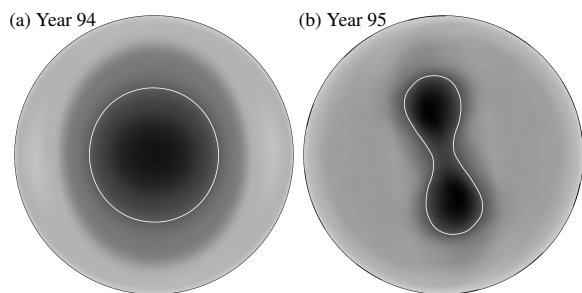


Figure 3. Winter mean potential vorticity over the North pole during the quiescent winter of year 94 (a) and the disturbed winter of year 95 (b) for $h_b = 0.08$. The white contour indicates the value $q = 0.6q_{\max}$, where q_{\max} is the maximum potential vorticity generated in an unforced integration ($h_b = 0$).

dynamically active disturbed winters. An alternative picture of the vortex structure and evolution is given by the potential vorticity field, $q = \zeta_a/h$. First, Figure 3 shows the winter mean potential vorticity field over the North pole averaged over the same 90-day period from the start of each year, once again for years 94 and 95 from the case with forcing amplitude $h_b = 0.08$. The level of potential vorticity with $q = 0.6q_{\max}$, where q_{\max} is the maximum value of obtained in a run with no wave forcing, is shown by the white contour and is broadly indicative of the vortex edge—this will be discussed further in section 3.2.2 below.

The potential vorticity distribution in year 94, Figure 3(a), is extremely simple, taking the form of an almost circular vortex at high latitudes, with a gradually more elliptical shape at low latitudes, a manifestation of the wavenumber-two forcing. There is a monotonic increase in values toward the pole. An examination of individual time frames throughout the winter (not shown) indicates a similarly simple evolution in time, with the pattern growing gradually in strength before decaying again.

In contrast, the mean potential vorticity in year 95, shown in Figure 3(b), is characterised by a distribution that is highly distorted from the circular equilibrium state. The mean vortex as defined by the white contour is almost deformed into two distinct lobes, while regions of higher potential vorticity are indeed split into two approximately circular lobes. As we have just seen, the vortex is highly transient in this winter and we do not expect such a deformed mean state to be maintained steadily in time.

Figure 4 shows individual snapshots of the potential vorticity for the winter of year 95 for this case. The twelve consecutive panels are not regularly spaced in time, but chosen over a 62 day interval to capture the key dynamical transitions. The figure shows that the vortex splits completely six times over the course of the winter, with each split followed by a partial reformation of the vortex into a monopolar, though still highly distorted distribution. The alignment of each split is approximately the same each time, being determined by the orientation of the topographic forcing, hence the time average shown in 3(b) captures much of the split vortex structure. The period of splitting and reformation is approximately 10 days but irregular and with intervals where the vortex remains split for more prolonged periods. The evolution is somewhat similar to the disturbed phases in the perpetual January investigations shown in Figure 3 of Scott (2016), and we conjecture that the dynamics here is in essentially the same regime, though restricted to a single winter at a time. Although the repeated splitting seen here is not found to the same extent in the winter stratosphere, the rapid reformation of the vortex following a split has occasionally been observed in the stratosphere (Scott 2016).

3.2. Vortex classification

In this section we quantify more carefully how the character of the interannual variability of the vortex changes across different wave forcing amplitudes. We consider two measures of the vortex state, based on the transience of the zonal velocity and the shape of the vortex (sections 3.2.1 and 3.2.2, respectively) and then consider the statistics of the distributions of quiescent and disturbed winters (section 3.2.3).

3.2.1. Vortex transience

A convenient diagnostic to quantify the vortex state is the winter mean transience of the zonal velocity field. This is defined as the winter mean of the magnitude of the zonal

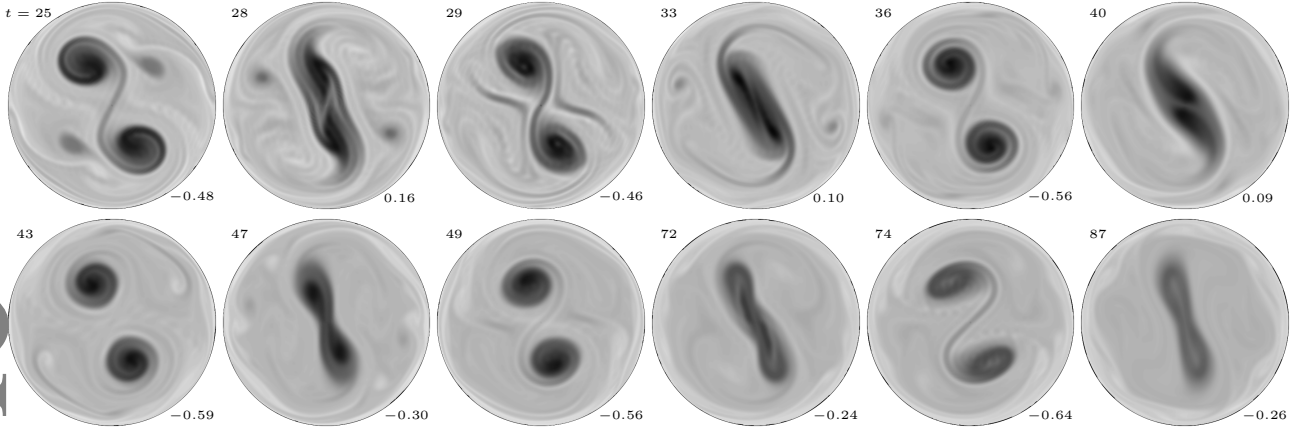


Figure 4. The evolution of the polar vortex during the 95th winter for $h_b = 0.08$ at selected times illustrating the vortex splitting and reformation. Times in days from the winter solstice indicated in the upper left; excess kurtosis shown for reference in lower right.

acceleration of the flow given by

$$\bar{T} = \frac{\int_{\phi_1}^{\phi_2} \int_{t_1}^{t_2} |\bar{u}_t| \cos \phi d\phi dt}{\int_{\phi_1}^{\phi_2} \int_{t_1}^{t_2} \cos \phi d\phi dt}, \quad (9)$$

where again $\phi_1 = 50^\circ$ and $\phi_2 = 70^\circ$, encompassing the approximate latitudinal extent of the vortex edge, and $t_1 = 0$ days and $t_2 = 90$ days from midwinter. The transience is essentially a measure of the wave-forcing of the vortex and the resulting unsteadiness of the zonal flow. In a quiescent year, when the vortex is increasing and decreasing gradually under the action of radiative forcing, the transience is typically around $1\text{--}2\text{ ms}^{-1}\text{day}^{-1}$. In a disturbed winter, in contrast, the rapid vortex splitting and reformation events result in a transience between about 6 and $9\text{ ms}^{-1}\text{day}^{-1}$.

With a single number as a characterization of the vortex state in a particular winter, we can conveniently summarize the interannual variability of the vortex across all forcing amplitudes. Figure 5 shows histograms for the winter mean transience for a selection of these. They demonstrate, in particular, unimodal distributions of vortex states at low and high wave forcing amplitudes, and an approximately bimodal distribution of vortex states at intermediate amplitudes.

The bimodality of the distributions can be quantified further by a *bimodality coefficient*, β , defined as (Ellison 1987)

$$\beta = \frac{g^2 + 1}{\kappa}, \quad (10)$$

where g is the discrete skewness of the distribution and κ is the discrete kurtosis of the distribution (not to be confused with the excess kurtosis of the vortex edge discussed in the following section). A perfectly unimodal Gaussian distribution has $\beta = \frac{1}{3}$, a perfectly uniform distribution has $\beta = \frac{5}{9}$, and a perfectly bimodal distribution (specifically, the sum of two distinct Dirac delta functions) has $\beta_e = 1$. Hence, we may consider a value of β greater than $\frac{5}{9} \approx 0.55$ as indicative of some degree of bimodality in the distribution. This quantification has limitations (and is non-unique) but is convenient as an approximate summary of the bimodality of each distribution.

The cases shown in Figure 5 are representative of three ranges of forcing amplitudes. At low forcing amplitudes, $h_b \leq 0.05$, every winter is quiescent, with a winter mean transience always lower than $3\text{ ms}^{-1}\text{day}^{-1}$. At the other end, at forcing amplitudes $h_b \geq 0.15$, every winter is disturbed with a winter mean transience always above $4\text{ ms}^{-1}\text{day}^{-1}$. The interval $3\text{--}4\text{ ms}^{-1}\text{day}^{-1}$ is a uniform minimum, containing very few winters across all forcing amplitudes, and marks a clear delineation between quiescent and disturbed winters. For intermediate forcing amplitudes lying in the range $0.065 \leq h_b \leq 0.125$, the bimodality of the distributions is large, with values always above $\beta = 0.7$. Note that the two cases shown in (b) and (c) are close to the end points of the bimodal range, with cases in between typically showing even stronger bimodality. In

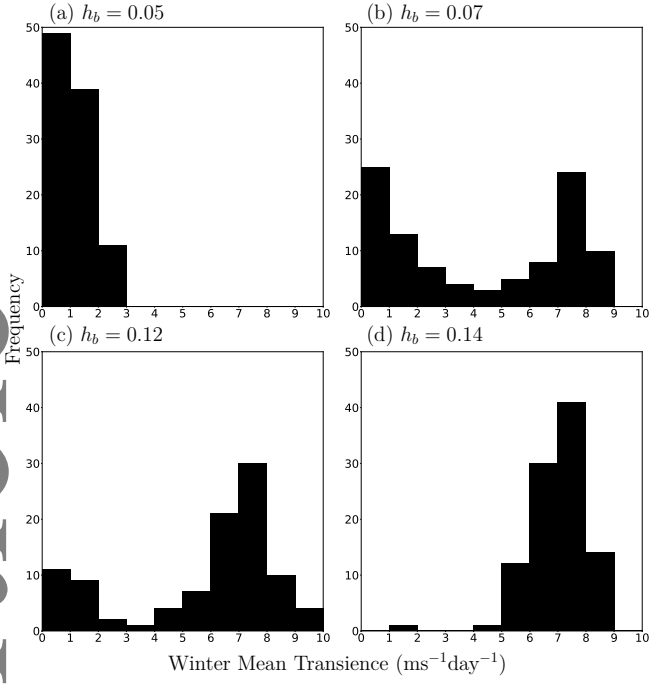


Figure 5. Histograms of latitude-weighted winter mean transience, as defined by (9). Top to bottom: $h_b = 0.05$, $h_b = 0.07$, $h_b = 0.12$ and $h_b = 0.14$.

all cases, the mean transience tends to fall into either the range below $3\text{ms}^{-1}\text{day}^{-1}$ or the range above $4\text{ms}^{-1}\text{day}^{-1}$, and with very few winters in the interval $3\text{--}4\text{ms}^{-1}\text{day}^{-1}$.

3.2.2. Potential vorticity morphology

We next consider the shape of the winter mean potential vorticity field, as in Fig. 3, across all winters. While the vortex edge at any given instant is most naturally defined as the region of steep potential vorticity gradients separating the vortex interior and the surf zone, the process of time averaging over the entire winter necessarily results in a blurring of potential vorticity gradients to the extent that the edge of the mean state is not always well-defined. For the purpose of our analysis, therefore, we define the edge of the winter mean vortex as the potential vorticity contour with value $q = 0.6q_{\text{max}}$, where q_{max} is the maximum potential vorticity generated in the unforced integration, $h_b = 0$ (white contour in Fig. 3). This value is somewhat arbitrary, as the vortex is growing and decaying over the course of the winter. However, visual inspection of potential vorticity fields at various times during the winter, and for a variety of forcing amplitudes, indicated that it gives a

reasonable approximation of the mean vortex edge in most cases.

Once the edge of the winter mean vortex is defined, we classify the winter state based on the vortex excess kurtosis, derived from the second elliptic moments as described in Scott (2016) and first used in this context in Matthewman and Esler (2011). Essentially, an excess kurtosis of zero corresponds to an elliptical vortex, positive values correspond to a deformation of an elliptical vortex with elongation at the tips, while negative values correspond to a deformation of an elliptical vortex by a pinching in the middle. The vortex shown in Fig. 3(b) is an example of the latter, with an excess kurtosis of approximately -0.3 .

After some experimentation, we found that it was optimal to classify as disturbed any winter whose mean vortex had an excess kurtosis below -0.05 . In fact, any small negative value was more or less acceptable, since in almost all quiescent winters the mean vortex excess kurtosis was consistently close to zero. It is worth noting that significant positive excess kurtosis was not recorded in any of the winters—which is to say the winter was never disturbed by virtue of filamentation on the outside edge of the vortex—hence there is no need to consider the threshold for positive excess kurtosis. Finally, as a measure to avoid misclassification of winters in cases of very large forcing amplitude, where it was occasionally found that the vortex was completely destroyed, an additional criterion was used that if the area of the winter vortex fell significantly below that of a typical disturbed winter then the winter was also classified as disturbed. These criteria were then applied to classify every winter mean vortex across all forcing amplitudes.

The time sequence of these winter states is shown in Figure 6 for various forcing amplitudes. As already noted, for the lowest forcing amplitudes, $0.04 \leq h_b \leq 0.055$, most if not all of the winters are quiescent and there is little or no interannual variability. For intermediate forcing amplitudes, $0.06 \leq h_b \leq 0.1$, there is an apparent regular oscillation between quiescent and disturbed winters, with a tendency

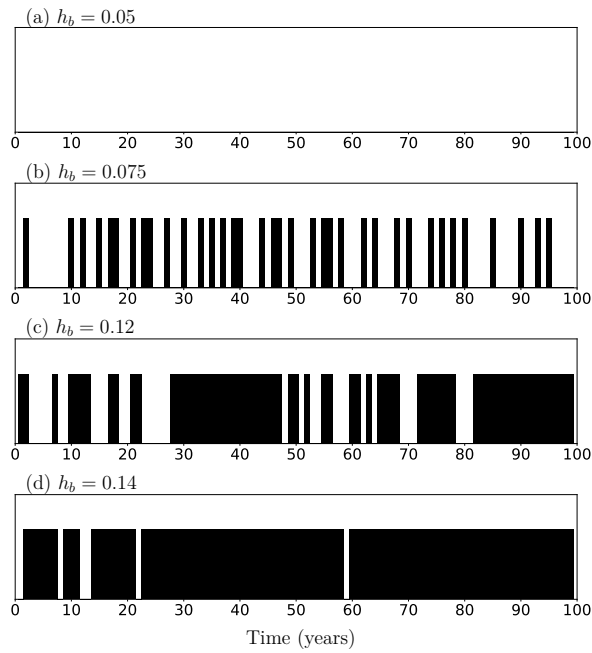


Figure 6. Sequences of quiescent and disturbed winters, as defined by the criteria described in the text, for low to high wave forcing amplitudes; black bars represent disturbed years. Top to bottom: $h_b = 0.05$, $h_b = 0.075$, $h_b = 0.12$ and $h_b = 0.14$.

for a quiescent winter to be immediately followed by a disturbed winter, and vice-versa. Indeed, this tendency is far greater than chance, as we shall see in section 3.2.3, and suggests that the flow conditions following a quiescent winter are such as to favour the development of a disturbed winter in the following year, and vice-versa.

Conversely, for the higher forcing amplitudes in the range $0.1 \leq h_b \leq 0.14$, we see the opposite behaviour. Instead of one kind of winter having a higher tendency to be followed by the other, winters have a tendency to come in strings of consecutively similar type, particularly disturbed ones, suggestive of a more regime-like behaviour. Again, this tendency appears far greater than chance as will be verified next. For example, in the case $h_b = 0.12$, it appears that there are two distinct regimes, one in which all winters are disturbed and another (e.g. between years 46 and 70) where disturbed and quiescent winters occurs with approximately equal frequency.

3.2.3. Wald-Wolfowitz Runs Analysis

To quantify the extent to which winter states are alternating, or forming regimes of consecutively similar states, we

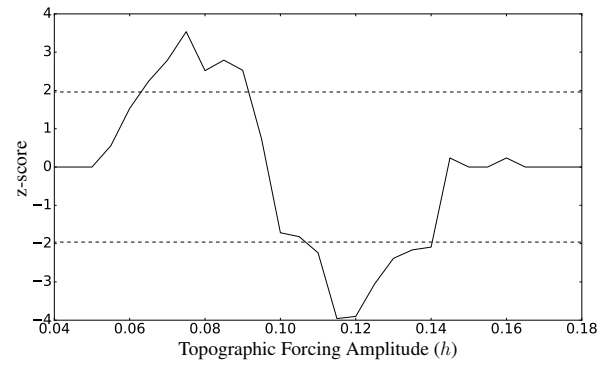


Figure 7. z -scores from Wald-Wolfowitz runs analysis vs forcing amplitude. The dashed lines are at $z = \pm 1.96$, beyond which are statistically significant results.

perform a Wald-Wolfowitz runs analysis (Bradley 1968). Essentially, this tests the null hypothesis that two types of event—here a quiescent winter and a disturbed winter—are mutually independent, and therefore that the distribution of a sequence of these events is random. If it is not random, the null hypothesis is rejected, and the events must be mutually dependent; in other words, a quiescent or disturbed winter depends on whether the previous winter or winters were quiescent or disturbed. This then provides information about the nature of the interannual variability of the system. A full description of the Wald-Wolfowitz runs test and its application to our experiments is given in appendix A.

We use the classification of quiescent and disturbed winters based on the potential vorticity field shape as defined in section 3.2.2 above. For each forcing amplitude we calculate a z -score, which is a measure of the dependence of each winter on previous winters. A positive z -score above $z = 1.96$ indicates that each winter is more likely than random to be of the opposite nature to the previous one with a statistical significance at the 95% confidence level. Conversely, a negative z -score below $z = -1.96$ indicates that each winter is more likely than random to be of the same nature as the previous one. As an example, the sequence obtained with $h_b = 0.075$, has a z -score of $z = 3.54$; the associated probability of occurring by chance is approximately 0.0002, sufficiently unlikely to reject the null hypothesis, and conclude that each winter does indeed depend on the preceding one.

The z -scores across the full range of forcing amplitudes considered are shown in Figure 7. Note that for very weak or very strong forcing amplitudes, $h_b \leq 0.05$ or $h_b \geq 0.165$, the sequences consist of entirely quiescent winters or entirely disturbed winters, respectively, and the test is not applicable. For intermediate forcing amplitudes in the range $0.065 \leq h_b \leq 0.09$, we have statistically significant positive z -scores, indicating a high dependence of the conditions of one winter on the previous one, with quiescent winters being significantly more likely to be immediately followed by disturbed winters, and vice-versa. This conclusion confirms the pattern inferred from Figure 6 and is consistent with the truncated primitive equation results of Scott and Haynes (1998).

Conversely, for higher forcing amplitudes in the range $0.11 \leq h_b \leq 0.14$ there are statistically significant negative z -scores, indicating a greater tendency for similar years to follow each other. In these cases, we note that there are also fewer quiescent winters than disturbed ones (e.g. Fig. 5), the asymmetry meaning that disturbed winters will tend to follow other disturbed winters. However, as we saw in Fig. 6(c) the distribution may be still more complicated, with long regimes in which all winters are disturbed, separated by regimes in which disturbed and quiescent winters occur with more equal likelihood. An analysis of individual subintervals would provide a more detailed picture of these latter regimes, but would require much longer time integrations to obtain a statistically significant result.

3.3. The Tropical Flywheel

As described in Scott and Haynes (1998), the subtropical flywheel is a mechanism whereby dynamical perturbations arising during one winter may persist until the following one owing to the weak effect of thermal relaxation on angular momentum anomalies at low latitudes. Since the variability found in the shallow water model at intermediate wave forcing amplitudes resembles closely that obtained in the truncated primitive equation model of Scott and Haynes

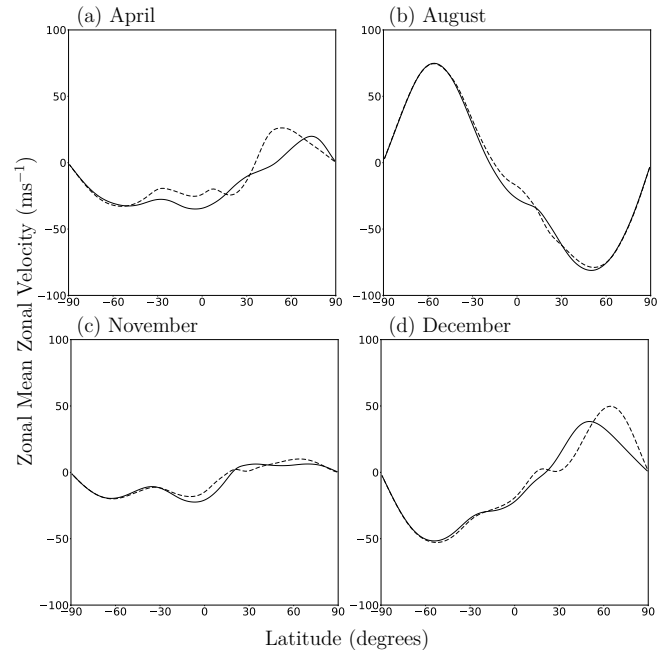


Figure 8. Monthly mean climatologies of the zonal mean zonal velocity for years preceding quiescent winters (solid line) and years preceding disturbed winters (dashed line) for $h_b = 0.075$: April, August, November, December (a–d).

(1998), we examine next the extent to which systematic flow anomalies are persistent at low latitudes from one year to the next.

Figure 8 shows composites of the monthly mean zonal mean zonal velocity for April, August, November, and December (model equivalent) for the case $h_b = 0.075$, the forcing amplitude which exhibited the strongest tendency for one type of winter to be followed by the other. The composites are separated into years which immediately precede a quiescent winter (solid line) and years which immediately precede a disturbed winter (dashed line). They therefore give a picture of the evolution of the zonal mean flow anomalies leading up to the two different kinds of winters.

Immediately after the winter, Figure 8(a), the zonal flow anomaly at mid to high latitudes is large, approximately 27ms^{-1} , since radiative damping has not had time to reduce the anomalies arising in the winter. Note that the sense of this anomaly is due to the fact that quiescent and disturbed winters are approximately alternating: thus, the solid line for pre-quiescent winters is comprised mostly of post-disturbed winters, and hence has stronger deceleration in

midlatitudes. In addition to the strong midlatitude anomaly, there is also an anomaly of approximately 15ms^{-1} across the tropics and subtropics. The anomaly has a dipolar structure, with the climatology of years preceding quiescent winters having stronger easterlies over the tropics and southern subtropics and weaker easterlies in the northern sub-tropics. In August, Figure 8(b), we can see that the anomaly has almost disappeared outside of the tropics, but remains persistent in the tropics, between approximately 25°S and 25°N . By December, Figure 8(c), the tropical anomaly has almost dissipated, but by this time it has already begun to influence the flow at higher latitudes for the following winter. We remark that the anomalies are relatively small compared with the differences they appear to control in the subsequent winter evolution.

A question now arises: if the tropical zonal wind anomaly is indeed responsible for the subsequent midlatitude winter evolution, might the strength and latitudinal extent of the friction implemented to control tropical easterlies have an influence on this control? To answer this, we conducted further experiments in which first the frictional timescale τ_{fr} and then the latitudinal extent of the friction ϕ_{fr} were varied. Considering first the dependence on τ_{fr} , Figure 9 shows the zonal mean zonal velocity for the final ten years of model integrations with various values of τ_{fr} . Figure 9(a) shows the case with an extremely weak drag $\tau_{\text{fr}} = 10000$ days. The tropical easterlies are uncontrolled, rapidly reaching physically unrealistic speeds and encroaching on the variability of the Northern winter polar vortex. Figures 9b and 9c show the cases $\tau_{\text{fr}} = 1000$ days and $\tau_{\text{fr}} = 100$ days respectively (timescales longer and shorter than the 365 days timescale used in section 3). It can be seen that increasing the damping has a marked effect on the behaviour of the winter evolution. For $\tau_{\text{fr}} = 1000$ days the majority of winters are disturbed and the vortex is slightly more confined to high latitudes than previously, due to the encroaching tropical easterlies, although even here there is still distinct and regular interannual variability of the type discussed above. For $\tau_{\text{fr}} = 100$ days the tropical easterlies

are much better controlled; now the majority of winters are quiescent, but these are punctuated by more isolated disturbed winters. For the extremely short timescale $\tau_{\text{fr}} = 10$ days (Fig. 9d), the tropical easterlies remain perpetually close to zero, but the interannual variability of the Northern winter polar vortex has also been completely quenched.

A similar pattern is found when increasing the latitudinal extent of the friction, with increasing ϕ_{fr} generally resulting in less distinct interannual variability (not shown). For example, with $\tau_{\text{fr}} = 100$, increasing ϕ_{fr} by a further 5° was enough to completely eliminate the variability seen in Figure 9(c) (in the sense that disturbed winters no longer appeared; small amplitude variability between quiescent winters persisted). However, the frictional timescale must be sufficiently large: with a one year timescale, increasing ϕ_{fr} even to 30°N was not enough to suppress the variability completely, the zonal flow anomalies being able to persist long enough in the tropics regardless of ϕ_{fr} .

So far we have been unable to determine more precisely what aspect of the tropical anomaly is responsible for determining the ensuing winter. Perturbation experiments were conducted in which a tropical force was applied in September, increasing the tropical winds preceding a quiescent winter and decreasing them preceding a disturbed winter. In neither case did the perturbed experiment show a change in the following winter. We speculate that not only the average tropical winds are important but also their latitudinal structure. In particular, there is a marked difference in the latitudinal shear across the subtropics, visible in Figure 8, that may also play a role in determining the subsequent evolution.

4. Modified radiative equilibrium

We return now to the method of controlling the tropical easterlies proposed in section 2.5: that of including a perturbation in the radiative equilibrium height profile to relax the flow in the tropics towards a westerly regime. This approach avoids the introduction additional forcing terms and does not directly interfere with momentum anomalies

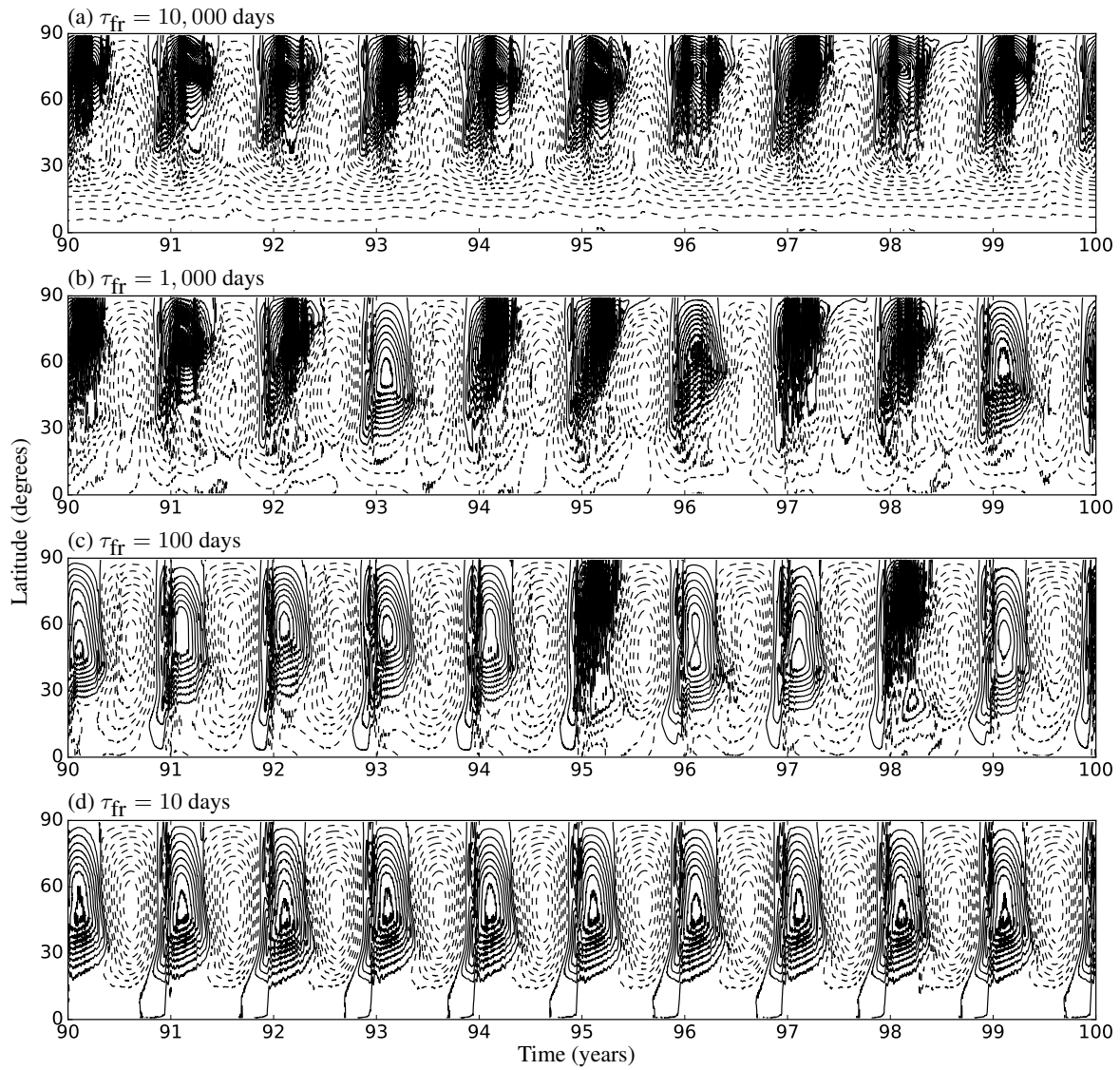


Figure 9. Latitude vs time contour plots of the zonal mean zonal velocity from the final 10 years of 100 year runs with $\phi_{fr} = 15^\circ$ and τ_{fr} set to various values. Contours as in Figure 1.

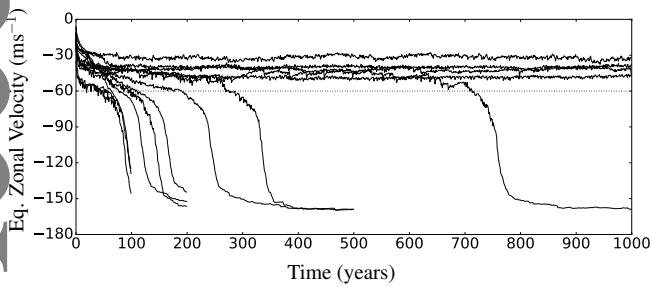


Figure 10. Equatorial zonal velocity for 1000 year time integrations using the additional local radiative equilibrium height profile over the tropics. The easterlies remain controlled for the full 1000 years (above the dotted line at -60ms^{-1}) for $0.06 \leq h_b \leq 0.075$. Reading left to right along the dotted line at -60ms^{-1} : $h_b = 0.095, 0.1, 0.04, 0.045, 0.09, 0.05, 0.055, 0.085, 0.08$. Note that the time series of runs that were not controlled for the full 1000 years were halted early.

latitude momentum means it might also be less effective at controlling the easterlies.

In fact, the method manages to control easterlies for the duration of the 100 year model integrations for intermediate wave forcing amplitudes in the range $0.05 \leq h_b \leq 0.09$. Outside this range, and most surprisingly, even for *lower* forcing amplitudes, all integrations produced a strong runaway easterly tropical flow.

In fact the collapse of the tropical flow to easterlies in this system appears to be rather sensitive to the forcing. When model integrations were carried for longer 1000 year durations, it was found that the range of forcing amplitudes able to constrain the easterlies over the full period was reduced to the much narrower range $0.06 \leq h_b \leq 0.075$.

that might be important in driving the variability. However, as discussed, the weak control of thermal relaxation on low

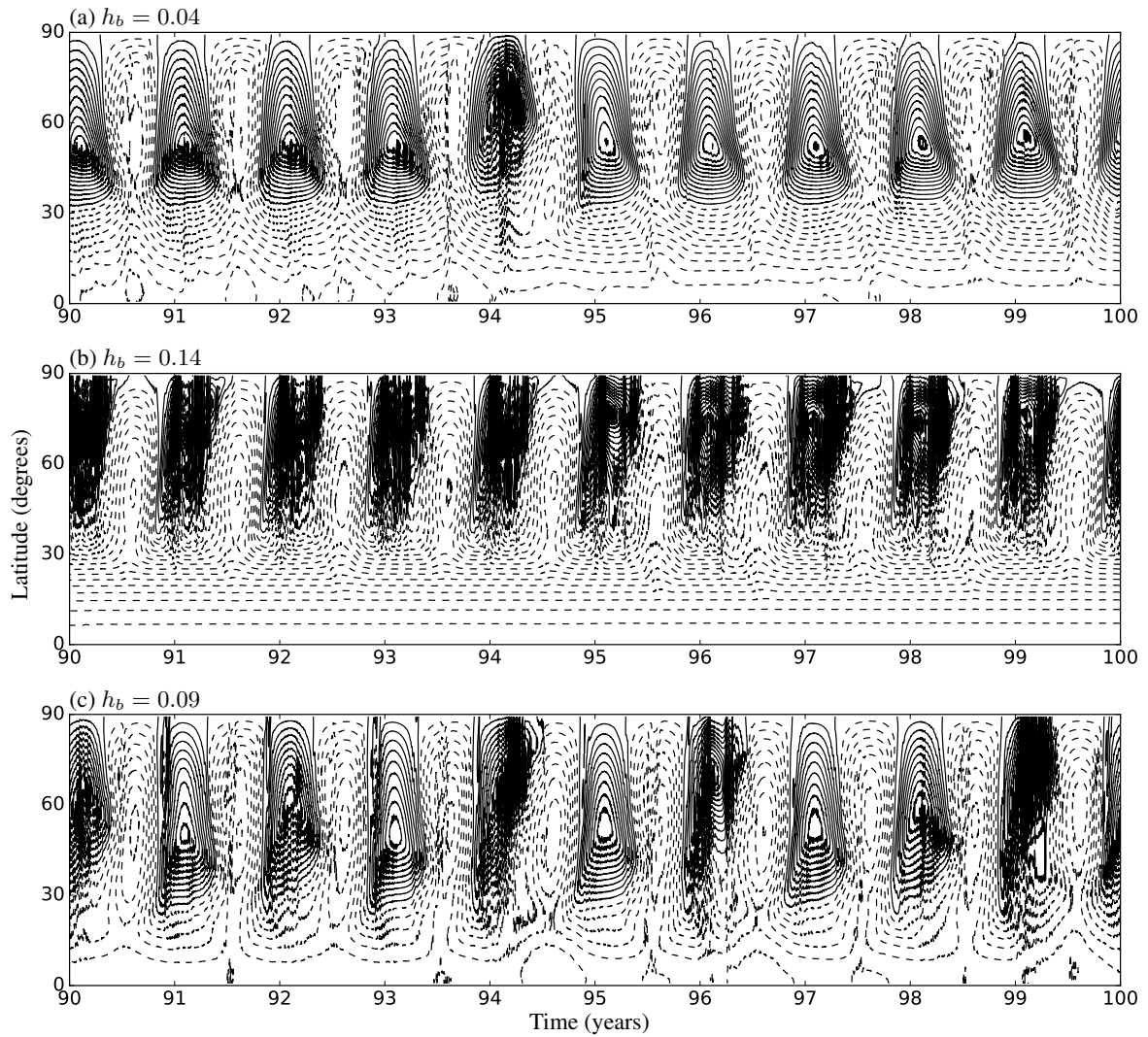


Figure 11. Zonal mean zonal velocity, \bar{u} , for the final 10 years of the radiative runs for various topographic forcing amplitudes (top to bottom: $h_b = 0.04$, $h_b = 0.12$ and $h_b = 0.09$). Contours as in Figure 1.

The dependence of the tropical flow (as measured on the equator) on forcing amplitude is shown in Fig. 10. It can be seen that integrations with forcing amplitudes outside this range show an abrupt collapse of the tropical flow: once the equatorial wind speed drops below about -60 ms^{-1} (dotted line) it tends to drop further very rapidly. The collapse happens sooner the greater the difference in forcing amplitude from the intermediate forcing range, both above and below. The case $h_b = 0.07$ appears to be the most stable forcing amplitude, remaining very steadily between approximately 20 and 40 ms^{-1} over the full 1000 years. Of course, it is still possible that even this and the other stable cases will eventually collapse if the integrations are continued long enough.

The collapse appears to be associated with the onset of an angular momentum conserving Hadley cell. This could be anticipated from the form of the radiative equilibrium modification, which consists of a local heating on the equator. The flow in thermal wind balance with such a heating has a local maximum on the equator and is unrealizable in a purely axisymmetric model (Held and Hou 1980). While the axisymmetric theory cannot be applied directly to a model with strong eddy activity, it nonetheless suggests there may be difficulties in attaining a state in thermal wind balance with the chosen radiative equilibrium state. Analysis of the zonal mean potential vorticity field (not shown) indicates that the circulation in the tropics in cases that exhibit the collapse to strong easterly flow is indeed in an angular momentum conserving state

(zero potential vorticity corresponding to constant angular momentum). However, the nonmonotonic dependence of the collapse time on forcing amplitude remains unknown. Since this is somewhat outside the scope of the present work we do not pursue it further here, and acknowledge that a single layer model is probably an inappropriate framework to study the interaction of an angular momentum conserving Hadley cell, a predominantly tropospheric phenomenon, and a stratospheric polar vortex.

Returning to the range of topographic forcing amplitudes for which the tropical easterlies are controlled over the original 100 year period, we once again find regular interannual variability, though of a slightly different nature from that described in section 3. Figure 11 shows the final ten years of integrations with low (a), high (b), and intermediate (c) forcing amplitude, showing clearly how the strong tropical easterlies in the first two encroach on the midlatitude flow, and the lack of interannual variability.

For forcing in the intermediate range, $h_b = 0.09$, interannual variability with a regular alternating pattern is again obtained, although the disruption to the polar vortex during disturbed winters is less pronounced than it was in the frictional runs. The variability in this case is better characterized by the latitudinal extent of the polar vortex, which, as can be seen in Figure 11(c), exhibits a quasi-regular variation of about 20° . To see this more clearly, we show in Figure 12 the latitude of the maximum of the winter mean zonal mean zonal velocity for all years in the 100 year integration. The regular oscillation of the latitude of the winter jet lasts the entire run.

Finally, Figure 13 shows the winter mean of the potential vorticity field over the North pole for years 55 and 56 of the same case, two typical successive winters. The difference between the two winters is less dramatic than that shown in Figure 3, but nevertheless shows a clear distinction, with the vortex significantly more elongated (left) in the year in which the jet maximum is further poleward.

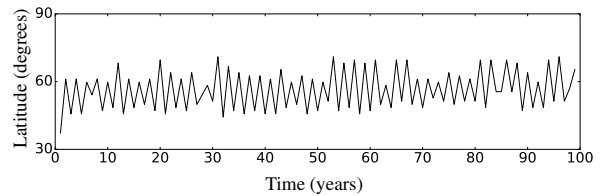


Figure 12. Evolution of the latitudinal maximum of the winter mean of the zonal mean zonal velocity for the $h_b = 0.09$ run.

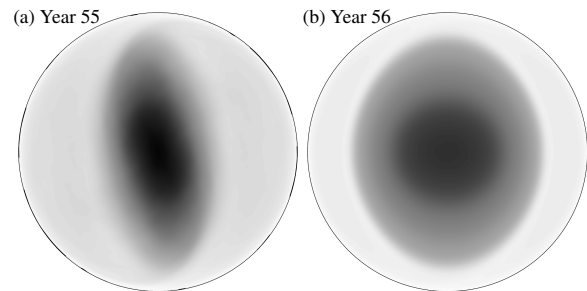


Figure 13. Winter mean of the potential vorticity field over the North pole for year 55 (a) and year 56 (b) of the $h_b = 0.09$ run.

5. Summary and Conclusions

Using a shallow water model with seasonally varying thermal forcing and time-independent topographic wave forcing, we have demonstrated the existence of coherent modes of internal low-frequency variability of the winter polar vortex. The vortex state in each winter is found to fall in one of two states, a quiescent state with low wave amplitudes and gradual changes to the mean flow following radiative equilibrium, and a disturbed state, with large wave amplitudes, a highly unsteady mean flow and large departures from radiative equilibrium. In the disturbed state the vortex is observed to split and reform repeatedly in a series of sudden warming-like events, though with more such events than typically observed during a single winter in the stratosphere. Relatively few winters fall into an intermediate, mildly disturbed category, though some may begin quietly and then transition to a disturbed state later than usual.

The low-frequency variability of the vortex occurs over a large range of wave forcing amplitudes and can be described in terms of the distribution of quiescent and disturbed winters. At intermediate topographic forcing amplitudes the vortex typically exhibits a quasi-regular sequence of alternating quiescent and disturbed winters: that is, quiescent

and disturbed winters are not distributed randomly, but have a statistically significant interdependence, with a quiescent winter more likely to follow a disturbed one and vice versa. At higher forcing amplitudes, the ratio of disturbed to quiescent winters increases and the variability exhibits regime-like behaviour, with long periods of purely disturbed winters distinct from periods in which the number of disturbed and quiescent winters is more evenly distributed.

The dynamical cause of the variability appears similar to the subtropical flywheel mechanism proposed by Scott and Haynes (1998), involving the weak constraint of radiative relaxation on the angular momentum at low latitudes. As in the stratosphere, summer easterlies prevent strong wave excitation by the topography, which leads to a summer circulation close to radiative equilibrium. Zonal mean flow anomalies at low latitudes, however, may persist throughout the summer and subsequently influence the evolution of the following winter. Specifically, enhanced mean flow deceleration following a disturbed winter leaves a predominantly negative flow anomaly throughout the tropics, although the latitudinal structure of the anomaly appears more complex, suggesting that the horizontal shear of the flow may also be important. At low latitudes the strength of the tropical/subtropical anomaly is approximately 10 ms^{-1} at the onset of the following winter.

The overall behaviour suggests a bistability of the winter vortex. For a given wave forcing, the vortex can do one of two things: it can enter into a quiescent state, coexisting stably with the wave forcing in such a way that the waves do not break and cause significant deceleration of the vortex; or it can enter into a transient, quasiperiodic state, repeatedly splitting and reforming. Which of these two states is selected appears to be determined in the early winter, while the vortex is forming, and appears to be quite sensitive to details of the flow in low latitudes. The behaviour is analogous to that found in the truncated primitive equation experiments of Scott and Haynes (2002). That it occurs here in a shallow water model indicates that the mechanism does not depend on the vertical structure of the vortex.

The behaviour found here is somewhat reminiscent of studies that have considered the correlation between the winter vortex and phase of the QBO in early winter and suggested that the winter evolution may be particularly sensitive to early winter QBO anomalies (Hampson and Haynes 2006; Anstey *et al.* 2010). However, the sense of the QBO influence on the vortex is in the opposite sense to the effect of the tropical anomalies found here, where an easterly tropical anomaly precedes a quiescent winter (Fig. 8). It is possible, therefore, that the full details of the structure of the anomalies across the tropics and subtropics needs to be taken into account. In Scott and Haynes (1998), for example, anomalies appeared to be centred off-equator, suggesting that latitudinal shear across the subtropics might be important, a pattern that was also seen in Figure 8. This, and the interplay of these anomalies with the QBO, is under current investigation.

The model integrations showing persistent long term internal interannual variability were made possible by including a weak frictional damping, with frictional timescale of one year, across the tropics. This was necessary to prevent the persistent increase of tropical easterly momentum in the model, as discussed in Scott (2016). The timescale of the friction is long enough that it did not unduly damp flow anomalies associated with the variability and when stronger damping was used the variability was suppressed. We note that the frictional nature of the low-latitude damping is unimportant. In further experiments in which the friction was replaced by a constant-in-time forcing, equal to the time average of frictional damping across the entire integration, almost identical results were obtained. Individual winters showed distinct evolution, as expected from a highly nonlinear system, but the number and distribution of disturbed winters was virtually unchanged. Such a fixed force may be interpreted as a crude parametrization of momentum fluxes from gravity wave breaking, with the one drawback that a complete

integration with frictional damping must be carried out first to determine the appropriate strength of the forcing.*

An alternative method of controlling the tropical easterlies, involving a simple modification to the radiative forcing in the tropics, was considered but found in general to be less effective. For intermediate forcing values the easterlies did indeed appear to be well constrained and interannual variability was obtained. However, much longer integrations revealed a destabilization of the tropical flow for all but a narrow range of forcing amplitudes, which appears to be due to the onset of angular momentum conserving flow in a Hadley cell. In cases where interannual variability was obtained with well-controlled tropical easterlies, it again took the form of an alternating series of winter states. In these cases, however, the difference between states was less extreme than in the cases with frictional damping, consisting of a shift in the location of maximum wind of the polar vortex and an alternation between vortices of more or less elliptical eccentricity. In almost no winters was the vortex observed to split. While highly speculative, it is conceivable that such quieter variability might be more relevant to the southern hemisphere winter vortex, where vortex splitting events are relatively rare, yet where there is evidence of distinct variability in the timing of the final vortex breakdown (Byrne and Shepherd 2018).

Appendix A: The Wald-Wolfowitz runs test

The Wald-Wolfowitz runs test calculates the number of “runs” of consecutive quiescent and disturbed winters expected under the null hypothesis that N_Q quiescent winters and N_D disturbed winters are distributed in a sequence at random. A “run” is a non-empty subsequence consisting of entirely of equal elements. For example, let ‘Q’ indicate a quiescent winter and ‘D’ a disturbed winter; then the following 12-element sequence: QQDDDDQQQQQD contains 4 runs, of length 2, 4, 5 and 1 in that order.

*The authors are grateful to an anonymous reviewer for suggesting the use of such a fixed forcing.

Under the null hypothesis of quiescent and disturbed winters being mutually independent events, the number of runs in the sequence of N winters containing N_Q quiescent winters and N_D disturbed winters, with $N = N_Q + N_D$, is a random variable with an approximately normal distribution, with a mean and variance given by (Bradley 1968)

$$\mu = \frac{2N_Q N_D}{N} + 1 \quad (11)$$

$$\sigma^2 = \frac{(\mu - 1)(\mu - 2)}{N - 1} \quad (12)$$

Note that these parameters do not assume that the number of quiescent and disturbed winters occur equally often, but only that they are independent and identically distributed. By classifying each winter as in section 3.2.2, we calculate the number of runs for each forcing amplitude sequence and compare it with the normal distribution predicted by the null hypothesis. If the number of runs is significantly higher or lower than the predicted mean, the null hypothesis is rejected, and the evolution of the flow during each winter is therefore to some extent dependent on the previous ones.

Using the sequence for $h_b = 0.08$ once again as an example, under our vortex classification criteria this sequence had 54 quiescent winters and 45 disturbed winters. The null hypothesis therefore predicts a normal distribution with mean and variance given by

$$\mu = \frac{2 \cdot 54 \cdot 45}{99} + 1 = 50.1 \text{ runs} \quad (13)$$

$$\sigma^2 = \frac{(50.1 - 1)(50.1 - 2)}{99 - 1} = 24.1 \text{ runs} \quad (14)$$

This also gives the standard deviation $\sigma = 4.9$. The actual number of runs in this model integration is 69, 3.85 standard deviations away from the mean predicted by the null hypothesis. This number of standard deviations away from the mean is called a standard score, or “z-score”.

From the z-score can be calculated the probability of obtaining it under the null hypothesis. If this probability is significantly low, the null hypothesis may be rejected. The probability of obtaining a z-score greater than or equal to a

given value x under the null hypothesis is given by

$$P(z \geq x) = \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{z}{\sqrt{2}} \right) \right), \quad (15)$$

where $\operatorname{erf}(z)$ is the error function of z defined by

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2) dt. \quad (16)$$

A value $z > 0$ indicates there were more runs than the null hypothesis predicted, and therefore that there was a greater tendency for quiescent winters to be immediately followed by disturbed winters, and vice-versa. Conversely, $z < 0$ indicates there were fewer runs than the null hypothesis predicted, and therefore that there was a greater tendency for the sequence to enter regimes of multiple quiescent or disturbed winters in a row. A z -score outside the range of $[-1.96, +1.96]$ has an associated probability of 0.025 or less and so is considered sufficiently statistically significant to reject the null hypothesis at the 95% confidence level.

References

- Anstey JA, Shepherd TG, Scinocca JF. 2010. Influence of the quasi-biennial oscillation on the extratropical winter stratosphere in an atmospheric general circulation model and in reanalysis data. *J. Atmos. Sci.* **67**: 1402–1419.
- Baldwin MP, Dunkerton TJ. 1999. Downward propagation of the arctic oscillation from the stratosphere to the troposphere. *J. Geophys. Res.* **104**: 30 937–30 946.
- Baldwin MP, Dunkerton TJ. 2001. Stratospheric harbingers of anomalous weather regimes. *Science* **294**: 581–584.
- Bradley J. 1968. *Distribution-free statistical tests*. Prentice-Hall.
- Byrne NJ, Shepherd TG. 2018. Seasonal persistence of circulation anomalies in the Southern Hemisphere stratosphere, and its implications for the troposphere. *J. Clim* **31**: 3467–3483.
- Chao WC. 1985. Sudden stratospheric warmings as catastrophes. *J. Atmos. Sci.* **42**: 1631–1646.
- Christiansen B. 1999. Stratospheric vacillations in a general circulation model. *J. Atmos. Sci.* **56**: 1858–1872.
- Ellison AM. 1987. Effect of seed dimorphism on the density-dependent dynamics of experimental populations of atriplex triangularis chenopodiaceae). *American Journal of Botany* **74**: 1280–1288.
- Esler JG, Scott RK. 2005. On the excitation of transient Rossby waves on the polar stratospheric vortex and the barotropic sudden warming. *J. Atmos. Sci.* **62**: 3661–3682.
- Gerber EP, Polvani LM. 2009. Stratosphere–troposphere coupling in a relatively simple AGCM: The importance of stratospheric variability. *J. Clim.* **22**: 1920–1933.
- Gerber EP, Voronin S, Polvani LM. 2008. Testing the annular mode autocorrelation time scale in simple atmospheric general circulation models. *Mon. Weather Rev.* **136**: 1523–1536.
- Hampson J, Haynes PH. 2006. Influence of the equatorial QBO on the extratropical stratosphere. *J. Atmos. Sci.* **63**: 936–951.
- Haynes PH. 2005. Stratospheric dynamics. *Ann. Rev. Fluid Mech.* **37**: 263–293.
- Held IM, Hou AY. 1980. Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *J Atmos Sci* **37**: 515–533.
- Holton JR, Dunkerton T. 1978. On the role of wave transience and dissipation in stratospheric mean flow vacillations. *J. Atmos. Sci.* **35**: 74–744.
- Holton JR, Mass C. 1976. Stratospheric vacillation cycles. *J. Atmos. Sci.* **33**: 2218–2225.
- Liu YS, Scott RK. 2015. The onset of the barotropic sudden warming in a global model. *Q. J. Roy. Meteor. Soc.* : doi: 10.1002/qj.2580.
- Matthewman NJ, Esler JG. 2011. Stratospheric Sudden Warmings as Self-Tuning Resonances. Part I: Vortex Splitting Events. *J. Atmos. Sci* **68**: 2481–2504.
- Matthewman NJ, Esler JG, Charlton-Perez AJ, Polvani LM. 2009. A new look at stratospheric sudden warmings. Part III. Polar vortex evolution and vertical structure. *J. Clim.* **22**: 1566–1585.
- MirRokni SM, Mohebalhojeh AR, Dritschel DG. 2011. Revisiting vacillations in shallow-water models of the stratosphere using potential-vorticity-based numerical algorithms. *J. Atmos. Sci.* **68**: 1007–1022.
- Rivier L, Loft R, Polvani LM. 2002. An efficient spectral dynamical core for distributed memory computers. *Mon. Weather Rev.* **130**: 1384–1390.
- Rong PP, Waugh DW. 2004. Vacillations in a shallow-water model of the stratosphere. *J. Atmos. Sci.* **61**: 1174–1185.
- Schoeberl MR, Strobel DF. 1980. Numerical simulation of sudden stratospheric warmings. *J. Atmos. Sci.* **37**: 214–236.
- Scott RK. 2016. A new class of vacillations of the stratospheric polar vortex. *Q. J. Roy. Meteorol. Soc.* **142**: 1948–1957.
- Scott RK, Haynes PH. 1998. Internal interannual variability of the extratropical stratospheric circulation: The low-latitude flywheel. *Q. J. Roy. Meteorol. Soc.* **124**: 2149–2173.
- Scott RK, Haynes PH. 2000. Internal vacillations in stratosphere-only models. *J. Atmos. Sci.* **57**: 2333–2350.

- Scott RK, Haynes PH. 2002. The seasonal cycle of planetary waves in the winter stratosphere. *J. Atmos. Sci.* **59**: 803–822.
- Scott RK, Polvani LM. 2006. Internal variability of the winter stratosphere. Part I: Time independent forcing. *J. Atmos. Sci.* **63**: 2758–2776.
- Scott RK, Rivier L, Loft R, Polvani LM. 2004. BOB: Model implementation and users guide. *NCAR Technical Note* : 30ppNCAR/TN-456+IA.
- Taguchi M, Yamaga T, Yoden S. 2001. Internal variability of the troposphere-stratosphere coupled system simulated in a simple global circulation model. *J. Atmos. Sci.* **58**: 3184–3203.
- Thompson DWJ, Baldwin MP, Wallace JM. 2002. Stratospheric connection to northern hemisphere wintertime weather: implications for prediction. *J. Climate* **15**: 1421–1428.
- Thompson DWJ, Wallace JM. 1998. The arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.* **25**: 1297–1300.
- Yoden S. 1987. Bifurcation properties of a stratospheric vacillation model. *J. Atmos. Sci.* **44**: 1723–1733.